

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 29, 2003, 15:16:07 ; Search time 40 Seconds
(Without alignments)
194.672 Million cell updates/sec

Title: US-09-924-102-2

Perfect score: 418
Sequence: 1 MLSTHFLFLFLFLFLSYSL.....RMGGGGRGGTADTGMEFLS 81

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1.*
2: p1r2.*
3: p1r3.*
4: p1r4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	72	17.2	845	2	D96799 hypothetical prote
2	71	17.0	627	4	A40201 artifact-warning s
3	68	16.3	462	2	H64048 signal recognition
4	68	16.3	511	2	A56750 archaen - human
5	67	16.0	118	2	T15239 hypothetical prote
6	67	16.0	247	2	S78470 NBDH2 dehydrogenas
7	67	16.0	829	2	S72366 DNA topoisomerase
8	67	16.0	1145	2	T18235 transcription acti
9	65	15.7	613	4	C40201 artifact-warning s
10	65	15.6	80	2	T10550 hypothetical prote
11	65	15.6	326	2	A41732 heterogeneous ribo
12	65	15.6	333	2	T52355 heterogeneous prote
13	65	15.6	386	1	S22315 snRNP-associated p
14	65	15.6	719	2	T52510 hypothetical prote
15	64.5	15.4	129	2	S14964 legumine-rich prote
16	64.5	15.4	3436	2	S55659 conserved hypotet
17	64	15.3	131	2	G90335 conserved hypotet
18	64	15.3	330	2	T25940 hypothetical prote
19	64	15.3	765	1	ISHU1 DNA topoisomerase
20	64	15.3	767	2	JU0144 DNA topoisomerase
21	64	15.3	767	2	A49546 DNA topoisomerase
22	64	15.3	1306	2	A70934 hypothetical glyci
23	63.5	15.2	389	2	T44957 heat shock protein
24	63.5	15.2	391	2	E84207 hypothetical glyci
25	63.5	15.2	749	2	A70812 hypothetical prote
26	63	15.1	163	2	T23076 hypothetical prote
27	63	15.1	432	2	B84620 hypothetical glyci
28	63	15.1	439	2	D70954 hypothetical glyci
29	63	15.1	476	2	AG1031 hypothetical glyci

30	63	15.1	1079	2	B70807 hypothetical glyci
31	62.5	15.0	276	2	T33925 hypothetical prote
32	62.5	15.0	463	2	T36810 probable integral
33	62.5	15.0	1410	1	A57013 early endosome ant
34	62	14.8	161	2	G71407 transcription fact
35	62	14.8	167	2	T20548 hypothetical prote
36	62	14.8	180	2	T28938 hypothetical prote
37	62	14.8	314	2	T08675 hypothetical prote
38	62	14.8	1898	1	A45973 trichhyalin - hum
39	61.5	14.7	268	2	S09860 hypothetical prote
40	61.5	14.7	272	2	T34164 hypothetical prote
41	61.5	14.7	284	2	T23158 hypothetical prote
42	61.5	14.7	806	2	S22765 heterogeneous ribo
43	61	14.6	78	2	B54897 dermaseptin b I pr
44	61	14.6	155	2	C86206 hypothetical prote
45	61	14.6	271	2	S34666 glycine-rich prote

ALIGNMENTS

```

RESULT 1
D96799
hypothetical protein F22K20.13 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: D96799
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Hutzlar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Lucos, J.S.; Maltl, R.; Marzla
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D96799
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-845 <STO>
A:Cross-references: GB:AE051173; NID:92829912; PIDN:AMC00620.1; GSPDB:GN00141
C:Genetics:
A:Gene: F22K20.13
A:Map position: 1
Query Match 17.2% Score 72; DB 2; Length 845;
Best Local Similarity 39.2% Pred. No. 6.9;
Matches 20; Conservative 9; Mismatches 18; Indels 4; Gaps 2;
Oy 28 LRKTKQOKKQEQIIRQSEVLEFRSETLRKTKGKKRRMGCGGGGCGTADTGG 77
Db 748 VRKEROOKANKVSYLSK---RGGRGGGRGARGGGANGRGSGSRDFGG 795
RESULT 2
A40201
artifact-warning sequence (translated ALU class A) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: A40201
R:Claverie, J.M.
personal communication, 1992
A:Reference number: A40201
A:Accession: A40201
A:Molecule type: DNA
A:Residues: 1-627 <CLA>
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potenti
A:Reference number: A40200; MUID:9241891; PMID:1572661
A:Contents: annotation

```

C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c in-frame stop codons are shown as 'x'.

C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

Query Match 17.0%; Score 71; DB 4; Length 627;

Best Local Similarity 44.4%; Pred. No. 6.7;

Matches 24; Conservative 4; Mismatches 24; Indels 2; Gaps 2;

OY 16 LSYSLGDRARLCL-RKTKOQKQEQILROSEVLFSEFSET-LRKTKGKRRMGQGG 67
DB 290 LHSLLDRAKRLKGGKXKXKXKXFFFLFFSEFSESRVAKAGVQWRDLG 343

RESULT 3

H64048 signal recognition particle 54K chain homolog fth - Haemophilus influenzae (strain Rd KW

C:Species: Haemophilus influenzae

C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 23-Jul-1999

C:Accession: H64048

R:Reischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.

; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.

; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.

Science 269, 496-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,

A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Reference number: A64000; M0ID:95350630; PMID:7542800

A:Accession: H64048

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1462 <TIGR>

A:Cross-references: GB:032696; GB:142023; M0ID:91573057; PIDN:AAC21784.1; PID:91573059; T

C:Genetics:

C:Superfamily: signal recognition particle 54K protein

Query Match 16.3%; Score 68; DB 2; Length 462;

Best Local Similarity 27.9%; Pred. No. 11;

Matches 29; Conservative 14; Mismatches 31; Indels 30; Gaps 5;

OY 3 LSTHFLFYFLSYSLGDRARL-----RLTKRKQKQEQILRO 43
DB 360 LSHVANKQYDDKRFVMEALINMTLKERANPDIIKSGRRRIALDSGVQDVNLLKQ 419

OY 44 SEVLFK-SETLKK-----TGKGRMGQGGGGGTADTGG 79
DB 420 FDEMQRMMKMRKGMKMGMLGGLGGLG---LGKMF 460

RESULT 4
A36750 archaean - human
C:Species: Homo sapiens (man)
C>Date: 08-Jul-1995 #sequence_revision 03-Aug-1995 #text_change 21-Jul-2000
C:Accession: A56750
R:Radice, P.; Pensotti, V.; Jones, C.; Perry, H.; Pierotti, M.A.; Tunnacliffe, A.
Genomics 26, 101-106, 1995
A:Title: The human archaean gene, ARCN1, has highly conserved homologs in rice and Drosop
A:Reference number: A56750; M0ID:95301274; PMID:7782067
A:Accession: A56750
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-511 <RAD>
C:Cross-references: GB:X81197; M0ID:9773572; PIDN:CAA57071.1; PID:9773573
C:Genetics:

A:Gene: GDB:ARCN1
A:Cross-references: GDB:377806; OMIM:600820
A:Map position: 11q23.3-11q23.3

Query Match 16.3%; Score 68; DB 2; Length 511;
Best Local Similarity 21.1%; Pred. No. 12;
Matches 20; Conservative 24; Mismatches 29; Indels 22; Gaps 2;

OY 3 LSTHFLFYFLSYSLGDRARL-----RLTKRKQKQEQILRO 42

DB 102 ISHCNCDLIFAFDEIVAGLVENVAQIRTFEMDSHEKVRVAVREOERAKAMRR 161

OY 43 QSEVLFSEFSETLRKTKGKRRMGQGGGGTADTGG 77

DB 162 KANEL--QOARRDAERQKRAKAPGFGGGSVAVSG 194

RESULT 5

T15239 hypothetical protein F56F4.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C:Accession: T15239

R:Murray, J.; Wohlmann, P.; Gilliam, B.

submitted to the EMBL Data Library, May 1997

A:Description: The sequence of C. elegans cosmid F56F4.

A:Reference number: Z18313

A:Accession: T15239

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-118 <MUR>

A:Cross-references: EMBL:AF03146; M0ID:92088763; PID:92088776; PIDN:AAB54204.1; GSPDB

A:Experimental source: strain Bristol N2; clone F56F4

A:Gene: CESP:F56F4.7

A:Map position: 1

A:introns: 58/1; 82/2

Query Match 16.0%; Score 67; DB 2; Length 118;

Best Local Similarity 29.4%; Pred. No. 3.7;

Matches 15; Conservative 13; Mismatches 21; Indels 2; Gaps 1;

OY 26 LCLRTKQKQEQILROSEVLFSEFSETLRKTKGKRRMGQGGGGTADTGG 76

DB 15 ICTTILVYVQRLKMPRAKRVKDKKGRKGRK--GKCKGSGSGTG 63

RESULT 6

S78470 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 6 - wheat mitochondrion

C:Species: mitochondrion Triticum aestivum (common wheat)

C>Date: 11-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 03-Jun-2002

C:Accession: S78470; S26224

R:Haouazine, N.; Pereira de Souza, A.; Jubler, M.F.; Lancelin, D.; Delcher, E.; Lejeu

Plant Mol. Biol. 20, 395-404, 1992

A:Title: The wheat mitochondrial genome contains an ORF showing sequence homology to

A:Reference number: S26224; M0ID:93043030; PMID:1421143

A:Accession: S78470

A:Molecule type: mRNA

A:Residues: 1-247 <HAO>

A:Cross-references: EMBL:X62100

A:Note: 27-Phe, 48-Phe, 67-Phe, 71-Phe, 72-Leu, 75-Tyr, 82-Leu, 167-Phe, and 208-Phe

A:Accession: S26224

A:Molecule type: DNA

A:Residues: 1-26, 'P', 28-47, 'P', 49-66, 'S', 68-70, 'SP', 73-74, 'H', 76-81, 'S', 83-166, 'S', 16

A:Cross-references: EMBL:X62100; M0ID:921825; PIDN:CAA44009.1; PID:921826

C:Genetics:

C:genome: mitochondrion

C:Superfamily: NADH dehydrogenase (ubiquinone) chain 6

C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation

Query Match 16.0%; Score 67; DB 2; Length 247;

Best Local Similarity 31.3%; Pred. No. 7.4;

Matches 26; Conservative 9; Mismatches 34; Indels 14; Gaps 4;

OY 3 LSTHFLFYFLSYSL-----GDRARLCLRTKQKQEQILROSEVLFSEFSETLRK 55

DB 158 LGLNLATYVFWFLVSSLLILVAMGAIYVTHRTK--VKRQDVFRNALDPRSHIMNR 215

OY 56 T-----GKGRMGQGGGGTAD 74

```

DB          216 TISPFHSHRR-SFSSGAGGPPD 237

RESULT 7
S72366
DNA topoisomerase (EC 5.99.1.2) I, somatic - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 12-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 16-Jul-1999
C:Accession: S72366
R:Pandit, S.D.; Richard, R.E.; Sternglanz, R.; Bogenhagen, D.F.
Nucleic Acids Res. 24, 3593-3600, 1996
A>Title: Cloning and characterization of the gene for the somatic form of DNA topoisomer
A:Reference number: S72366; PMID:96433160; PMID:8836188
A:Accession: S72366
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-829 <SPAN>
A:Cross-references: EMBL:L07777; NID:g214833; PIDN:AAB36608.1; PID:g214834
A:Experimental source: somatic cells
C:Comment: The type I DNA topoisomerase catalyzes the ATP-independent transient breakage
reack in another, followed by rejoining. This reaction will lead to the conversion of one
C:Genetics:
A:Gene: TOP1
C:Superfamily: eukaryotic type I DNA topoisomerase
C:Keywords: DNA binding; DNA replication; isomerase
F:7/5/Active site: Tyr #status predicted

Query Match      16.0%; Score 67; DB 2; Length 829;
Best Local Similarity 30.0%; Pred. No. 24;
Matches 18; Conservative 9; Mismatches 33; Indels 0; Gaps 0;

OY      18 YSLGRARLCRLKTKQOKKEQQLROSEVLFRSTLRRTGGKRRWGQGGRGTADTGC 77
        | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      500 YETARLKMCVEKIRMTYKEDMKSKEMKVQRALVALYFDIKLAIRAGNEKEGETATDTVG 559

RESULT 8
T18235
transcription activator GAL11 homolog - yeast (Candida albicans)
C:Species: Candida albicans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18235
R:Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, November 1998
A:Reference number: Z18831
A:Accession: T18235
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1145 <BAR>
A:Cross-references: EMBL:AL033501; NID:e1341022; PID:e1341031; PIDN:CAA21993.1
C:Genetics:
A>Note: CA4IC10.09

Query Match      16.0%; Score 67; DB 2; Length 1145;
Best Local Similarity 30.4%; Pred. No. 32;
Matches 21; Conservative 15; Mismatches 23; Indels 10; Gaps 3;

OY      16 LSYSGDRARLCRLKTKQOKKEQQLROSEVLFPSR--TLRKTRKKRRWGGGGR----- 69
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      842 IOHPMSGPQQVLVQDDQDDQDDPPOHQSVSKASQSVAKTNKKT--GQGRKKKASI 898
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY      70 -GGTADTGC 77
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      899 SAGTAPTPG 907

RESULT 9
CA0201
artifact-warning sequence (translated ALU class C) - human
C:Species: Homo sapiens (man)
C>Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: CA0201
C:Clavette, J.M.
```

```

personal communication, 1992
A:Reference number: A40201
A:Accession: C40201
A:Molecule type: DNA
A:Residues: 1-613 <CLA>
R:Glover, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Ali-derived and other potentially
A:Reference number: A40200; MUID:92241891; PMID:1572661
A:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames c
in-frame stop codons are shown as 'x'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of

Query Match          15.7%; Score 65.5; DB 4; Length 613;
Best Local Similarity 30.1%; Pred. No. 26;
Matches 22; Conservative 5; Mismatches 15; Indels 31; Gaps 3;

OY 16 LSYSLGDRARLCLRTKTKQOKEQILRQSEVLFNSETLRKTKGK-----RRMGQG 67
DB 184 LHSSIGDRVRLCLKKKK-----KKXXXGVAHVACNPSTLGGKG 223
      ||||| ||||| :| :| :|
OY 68 G---RGATDTGC 77
DB 224 GWMRPVGRDPG 236
      | | | | |
      | | | | |

RESULT 10
T10550
hypothetical protein T12G13.70 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Oct-1999
C:Accession: T10550
R:Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bam
submitted to the Protein Sequence Database, June 1999
A:Reference number: Z16533
A:Accession: T10550
A:Molecule type: DNA
A:Residues: 1-80 <BEV>
A:Cross-references: EMBL:AL080252; GSPDB:GN00062; ATSP:T12G13.70
A:Experimental source: cultivar Columbia; BAC clone T12G13
C:Genetics:
A:Gene: ATSP:T12G13.70
A:Map position: 4

Query March          15.6%; Score 65; DB 2; Length 80;
Best Local Similarity 41.9%; Pred. No. 4.2;
Matches 13; Conservative 5; Mismatches 11; Indels 2; Gaps 1;

OY 49 RSETLRKTGKKRRMG--OGGRCGTADTGC 77
DB 3 KSDAYRRKSGKPKNKKMGGMGGGGGGSGG 33
      :| :| :| :| :| :| :| :| :| :|
      :| :| :| :| :| :| :| :| :| :|

RESULT 11
A41732
heterogeneous ribonuclear particle protein hrp36 - fruit fly (Drosophila melanogaster)
N:Alternate names: heterogeneous nuclear RNP protein hrp36
C:Species: Drosophila melanogaster
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 23-Jul-1999
R:Matunis, E.L.; Matunis, M.J.; Dreyfuss, G.
J. Cell Biol. 116, 257-269, 1992
A:Title: Characterization of the major hnRNP proteins from Drosophila melanogaster.
A:Reference number: A41732; MUID:92112966; PMID:1730754
A:Accession: A41732
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-326 <MAT>
A:Cross-references: GB:K62636; NID:g11035; PTDN:CA44502.1; PTD:g11036
A:Note: Sequence extracted from NCBI Backbone (NCBIN:76634, NCBI:P:76635)
C:Genetics:

```


	Matches	14;	Conservative	3;	Mismatches	6;	Indels	1;	Gaps	1;
Oy	57	GKGRBWMGO-GGGRGTADTGWF	79							
Db	52	GGGGRGTGKSGGGRGGRDSSGRF	75							

Search completed: May 29, 2003, 15:21:03
 Job time : 42 secs

THIS PAGE BLANK (USPTO)